Claims

What is claimed is:

1. An apparatus for facilitating data clustering, said apparatus comprising:

an arrangement for obtaining input data; and

5   an arrangement for creating a predetermined number of non-overlapping subsets of the input data;

said arrangement for creating a predetermined number of non-overlapping subsets being adapted to split the input data recursively.

2. The apparatus according to Claim 1, wherein said arrangement for creating a

10   predetermined number of non-overlapping subsets is adapted to initially split the input data into at least two sets of output data.

3. The apparatus according to Claim 2, wherein said arrangement for creating a predetermined number of non-overlapping subsets is adapted to:

split the at least two sets of output data recursively; and

repeat the recursive splitting of output data sets until the predetermined number of non-overlapping subsets is obtained.

4. The apparatus according to Claim 2, wherein said arrangement for creating a predetermined number of non-overlapping subsets is adapted to determine an eigenvector decomposition relating to the input data.

5. The apparatus according to Claim 4, wherein said arrangement for creating a predetermined number of non-overlapping subsets is adapted to determine a vector of projection coefficients onto the set of eigenvectors in the eigenvector decomposition.

6. The apparatus according to Claim 5, wherein said arrangement for creating a predetermined number of non-overlapping subsets is adapted to determine a probability density relating to the vector of projection coefficients.

7. The apparatus according to Claim 6, wherein said arrangement for creating a predetermined number of non-overlapping subsets is adapted to:

assign at least one threshold relating to the probability density; and

yield the at least two sets of output data based on the relation to the threshold of a value associated with a function relating to the projection coefficients.

8. The apparatus according to Claim 7, wherein there are N-1 thresholds, where N is the number of sets of output data to be yielded.

9. The apparatus according to Claim 8, wherein each threshold is a value of the function relating to the projection coefficients for which the probability density equals m/N, where m is a number from 1 to N-1.

10. The apparatus according to Claim 1, wherein the data clustering relates to the enrollment of target speakers in a speaker verification system.

11. A method of facilitating data clustering, said method comprising the steps of:

obtaining input data; and

creating a predetermined number of non-overlapping subsets of the input data;

step of creating a predetermined number of non-overlapping subsets comprising splitting the input data recursively.

12. The method according to Claim 11, wherein said splitting step comprises initially splitting the input data into at least two sets of output data.

13. The method according to Claim 12, wherein said splitting step comprises:

splitting the at least two sets of output data recursively; and

repeating the recursive splitting of output data sets until the predetermined number of non-overlapping subsets is obtained.

14. The method according to Claim 12, wherein said splitting step comprises determining an eigenvector decomposition relating to the input data.

15. The method according to Claim 14, wherein said splitting step further comprises determining a vector of projection coefficients onto the set of eigenvectors in the eigenvector decomposition.

16. The method according to Claim 15, wherein said splitting step further comprises determining a probability density relating to the vector of projection coefficients.

17. The method according to Claim 16, wherein said splitting step further comprises:

assigning at least one threshold relating to the probability density; and

yielding the at least two sets of output data based on the relation to the threshold of a value associated with a function relating to the projection coefficients.

18. The method according to Claim 17, wherein there are N-1 thresholds, where N is the number of sets of output data to be yielded.

19. The method according to Claim 18, wherein each threshold is a value of the function relating to the projection coefficients for which the probability density equals m/N, where m is a number from 1 to N-1.

20. The method according to Claim 1, wherein the data clustering relates to the enrollment of target speakers in a speaker verification system.

21. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for facilitating data clustering, said method comprising the steps of:

obtaining input data; and

creating a predetermined number of non-overlapping subsets of the input data;

step of creating a predetermined number of non-overlapping subsets comprising

splitting the input data recursively.